

◆人工智能算法与应用◆

基于过采样和卫星图像多特征融合的街区品质评估^{*}郭茂祖^{1,2,*}, 王偲佳^{1,2}, 王鹏跃^{2,3}, 李 阳^{1,2}, 赵玲玲⁴

(1. 北京建筑大学电气与信息工程学院, 北京 100044; 2. 建筑大数据智能处理方法研究北京重点实验室(北京建筑大学), 北京 100044; 3. 北京建筑大学建筑与城市规划学院, 北京 100044; 4. 哈尔滨工业大学计算机科学与技术学院, 黑龙江哈尔滨 150001)

摘要:城市街区的品质评估是城市规划中的重要步骤之一。针对已有街区品质评估模型中使用的街景图像采集位置和角度不统一造成的图像内容偏差问题, 本文提出使用视角统一且易获取的卫星图像代替街景图像。同时考虑多种街区品质测度指标, 将非图像型数据与图像型数据进行特征融合, 从多角度进行城市街区品质表达, 然后针对小样本和数据不平衡场景, 结合机器学习中的随机过采样和随机森林(Random Forest, RF), 基于融合数据进行城市街区的品质评估。实验结果表明, 本研究方法可以有效进行全面的城市街区品质表达, 并有效降低过拟合, 对比未使用过采样和多特征融合的方法, 正确率提高约8个百分点。

关键词:城市街区品质评估 特征融合 朴素随机过采样 随机森林 卫星图像

中图分类号: TP391.41, TP18 文献标识码: A 文章编号: 1005-9164(2022)04-0652-08

DOI: 10.13656/j.cnki.gxkx.20220919.005

街区作为城市结构的基本组成部分, 是城市规划和设计中的重要研究对象。城市街区的品质、功能等方面的评估分析工作也是城市规划和设计工作中必不可少的一环。城市规划和设计领域中所定义的街区通常是指被道路包围的区域, 或借助其他自然特征或人文特征进行划分的区域。街区的品质评价方法能够以统一的测度指标对各街区的品质进行客观评估。街区品质评估的结果可以作为街区品质提升工作和后期街区建设规划工作的基础, 有着重要的价值

和意义。

已有的城市街区品质评估方法侧重于物质空间及社会属性^[1]。目前对于城市街区品质评估的研究, 以及街区品质评估体系的构建、测度指标的制定和品质类别的区分等仍以定性研究^[2,3]为主。街区品质类别的确定主要由业内资深专家基于业内文字概念的定义、描述以及从业多年的经验积累进行主观判断, 由于不同职业培养体系和不同文化背景下的业内专家对于城市街区品质类别的认知和偏好略有不同, 所以导致主观判断得出的街区品质评估结果不统一,

收稿日期: 2022-04-30

* 国家自然科学基金面上项目(61871020, 62101022), 北京建筑大学市属高校基本科研业务费专项(X20161)和北京市属高校高水平创新团队建设计划项目(IDHT20190506)资助。

【作者简介】

郭茂祖(1966-), 男, 教授, 博士生导师, 主要从事机器学习、智能建造、智慧城市和生物信息学研究, E-mail: guomaozu@bucea.edu.cn。

【* * 通信作者】

【引用本文】

郭茂祖, 王偲佳, 王鹏跃, 等. 基于过采样和卫星图像多特征融合的街区品质评估[J]. 广西科学, 2022, 29(4): 652-659.

GUO M Z, WANG S J, WANG P Y, et al. Block Quality Evaluation Based on Oversampling and Satellite Image Multi-feature Fusion [J]. Guangxi Sciences, 2022, 29(4): 652-659.

缺乏科学的评判标准。

针对这些弊端,有研究人员提出了基于街景的智能评价方法^[4,5]和基于生物传感器的感知评价方法^[6]。在近几年的研究中,Rundle等^[7]利用街景图像对纽约37个步行街区环境美感与其他物质空间指标进行评价。Naik等^[8]通过街景图像评价美国21个城市街道空间的安全度。Ewing等^[9]基于已有的指标体系从不同维度对街景图像的空间围合度、意象性等量化评价进行打分。韩君伟^[10]借助数字摄影和计算机图像处理技术将视觉熵引入量化商业步行街道景观视觉复杂性的研究中。龙瀛等^[11]将大数据与街景数据结合,对北京和成都的街道空间进行品质分析,并得出评价指标为人口密度、城市活力、界面特征、交通特征、开发强度、可达性等。唐婧娴等^[4]通过计算机图像识别、图像分割技术和人口访问调查方法研究北京和上海的中心区街道品质,得出街道环境设计要素为绿化率、街道高宽比、街道尺度、街道活力、街道围合度、人性化尺度、通透性、整洁度、意象化等。樊钧等^[12]在对苏州古城区街道空间慢行品质的研究中,通过空间网络分析、聚类算法等技术分析,得出评价指标为兴趣点、位置服务数据、街道尺度、组织结构、绿化率、空间活力、设施均好性、街道五感等。这些方法合理降低了专家评估的主观性,但因为只侧重于某测度指标中的某一方面,所以导致评估所参考的指标不够全面。

随着开源社区的发展,城市中累积的海量多源大数据为城市街区的研究提供了更宽广的角度和思考,其中各街区品质测度指标数据主要包括实地调查数据^[13,14]、地理信息数据^[15,16]、街景图片数据^[17-19]、空间句法评价^[20,21]等。

人工智能的飞速发展促使各领域研究人员将人工智能用于解决行业内的部分问题。在城市规划领域中,过去专家常用人工方式评估街区品质,但随着数据逐渐增多,人工评估的时间成本和人力成本增大。另外,专家的从业经验和个人观念使得街区品质评估带有主观性,从而无法严格执行现有规范。基于人工智能的街区品质评估方法可有效缓解上述问题,但目前使用人工智能的街区品质评估方法与实际应用的研究成果较少。本文综合研究人员在各方面提出的街区品质测度指标,提出全面的城市街区品质测度指标,并综合城市街区品质评估的研究情况,提出使用卫星图像取代街景图像作为输入图像,降低街景图像采集位置与视角不统一造成的偏差;将非图像型

数据与图像型数据提取的特征进行融合,从多角度充分提取街区的品质特征;针对小样本和数据不平衡场景,本文使用朴素随机过采样方法削弱真实数据集的不平衡性对模型训练的影响,并使用随机森林(Random Forest,RF)作为街区品质类别的分类器,以提高街区品质评估模型的适用性和泛化性。

1 相关工作

根据不同街区相关数据得到的特征,可分为非图像型数据和图像型数据,其中非图像型数据以各项人工采集或传感器采集的数值为主,如研究对象的数目或比值等,各项数值组合在一起形成一维特征向量,通过计算特征向量在多维度坐标中的距离等进行分析从而完成街区品质评估;图像型数据以彩色照片或灰度图为主,经由图像特征提取得到特征矩阵,并经过特征矩阵间的计算完成街区品质评估。虽然非图像型数据与图像型数据在格式和计算方法上略有区别,但步骤和原理相近。

图像型数据除像素矩阵形式外,还可基于图像识别、语义分割等方法通过识别图像中对象的种类数、数目、面积占比和在图像中的位置等进行信息提取,并将得到的图像特征与数值特征进行特征融合。特征融合问题常见于多模态的研究中,较为常见的特征融合方法主要为拼接和对位求和。

针对真实数据集的不平衡性易导致提取到的特征向量类别不平衡,从而影响训练效果的问题,有关研究常对数据集进行过采样和欠采样等处理。常见的过采样方法为朴素随机过采样、合成少数类过采样技术(Synthetic Minority Over-sampling Technique, SMOTE)^[22]、合成边缘少数类过采样技术(Borderline-Synthetic Minority Over-sampling Technique, Borderline-SMOTE)^[23]、自适应合成抽样(Adaptive Synthetic Sampling, ADASYN)^[24]等,欠采样方法为朴素随机欠采样等。由于真实数据集的街区品质类别标签须由人工标注,人力、物力成本高,所以数据量较少。朴素随机过采样方法利用有放回的抽取方法,从少数类样本中随机采样,以此代替现有的数据集,其优点在于增加了少数类样本的权重,但有可能造成过拟合。本文选择朴素随机过采样作为城市街区品质评估中的预处理方法。

城市街区的品质评估问题可转化为分类问题,其中专家标注的街区品质分为A、B、C 3个类别。真实数据集包含141个街区的数据,属于小型样本,且以

非图像型数据为主,故机器学习算法更适合用于以上问题。基于机器学习的分类算法以支持向量机(Support Vector Machine, SVM)、决策树、随机森林为主。

2 城市街区品质评估方法

本文使用卫星图像,通过指定颜色范围的内容识

别与占比计算,提取各街区的绿化比率值,并与非图像型特征进行特征融合。融合后的特征进行朴素随机过采样处理,减少数据的不平衡性对模型训练的影响。模型使用随机森林对城市街区进行分类,通过品质类别的确定完成城市街区的品质评估,具体结构如图1所示。

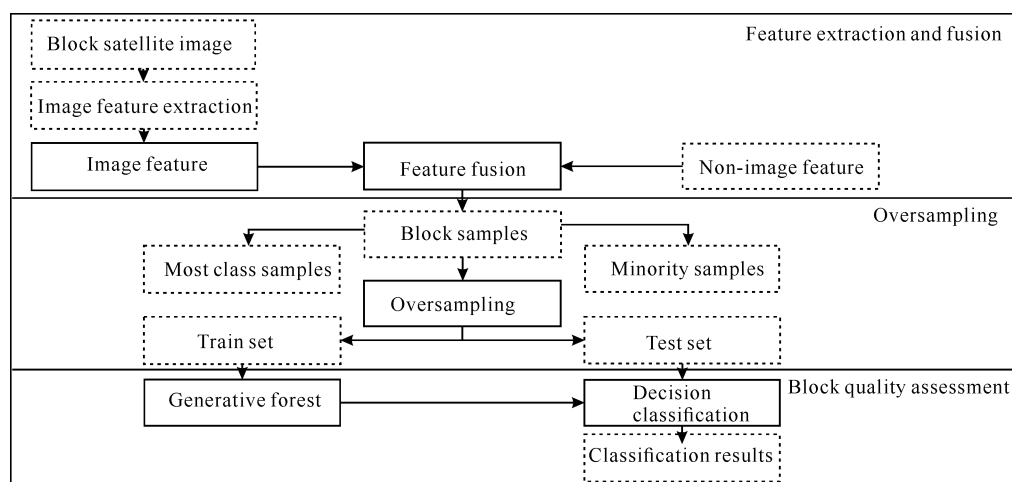


图1 街区品质评估模型结构

Fig. 1 Model structure of block quality evaluation

2.1 基本定义

为全面、客观、精确地对影响城市街区各品质评估的因素进行表达,本文以物质空间实地调研、空间句法效能分析以及开源数据获取等方式构建城市街区品质的测度指标:建筑风貌、重要历史建筑数量、街道建筑高宽比、沿街建筑功能种类、小品服务设施种类、停车位数量、公交地铁站点站距、停车干扰度、人行道宽度、沿街空间出入口数量、绿化覆盖率、植被丰富度、道路交通可达性、街道交通潜力、沿街开敞空间可达性、空间视线吸引力、公共交通步行可达性、公共基础服务设施可达性、人流聚集度、功能密度、天空宽阔度、功能混合度和绿化占比,其中,绿化占比是由图像提取的特征,简称为图像型特征。以上指标参考近年来城市规划专家关于街区品质评估问题所提出的评价体系与相关指标^[25-29],从自然环境、人文环境和社会环境等方面综合评价街区公共空间的品质。

在城市街区品质测度指标中,由卫星图像计算得出的绿化占比与由街景图像计算得出的绿视率在定义上相近,二者在街区绿化品质的表达上存在互补作用。街景图像是以人类视角在街道某一特定点位,以街道中线为基准拍摄的图像,主要由中间道路、两侧的绿化和建筑等设施组成。卫星图像的拍摄视角为

俯视,其拍摄到的绿化情况与侧视拍到的街景图像绿化情况不同。街景图像中的绿视率是以人为本的感官品质指标,而绿化占比注重公共空间内客观存在的绿化面积,两者互补。故本文提出一种街区品质评估的改进方法,即利用卫星图像代替街景图像,并将图像特征与数值特征融合,基于随机森林构建街区品质的分类模型。

2.2 特征提取与融合

首先将由物质空间实地调研、空间句法效能分析以及开源数据等方式获取到的街区非图像型数据进行整理,验证是否存在缺省值并将缺省特征补零。

2.2.1 街区卫星图像分割与预处理

通过谷歌卫星地图获取高清卫星图像,并结合街区轮廓矢量数据进行框选和截取,获得严格按照街区轮廓划分的卫星图像并编号。由于模型输入图像须为规则图形,故将原街区图像取最小外接矩形并全白填充背景,避免无关内容影响模型训练。此外,本文预先对截取的街区图像进行锐化处理,增强图像中各边界、轮廓线和其他细节,使其更加清晰。

2.2.2 街区卫星图像特征表示

图像型特征主要通过图像矩阵中各像素值的计算进行提取,图像型数据集中的街区卫星图为 RGB

三通道图像,通过各通道间同一位置像素值的比较,确定此位置是否属于绿化范围,并计算代表街区绿化特征的像素数在整体街区像素数中的占比,即街区绿化占比。绿化占比作为样本的图像型特征,以首尾拼接的方式与多维数值特征进行融合。

2.3 过采样

使用朴素随机过采样方法对融合后的特征进行预处理,并与 SMOTE、Borderline-SMOTE 和 ADA-SYN 等常见的过采样方法进行对比。

2.4 基于随机森林的分类器模型

本文使用随机森林算法作为分类器,随机森林是一种特殊的使用决策树作为模型的 bagging。首先,用 bootstrap 方法生成多个训练集,并分别对每个训练集构造一棵决策树,然后在节点寻找特征进行分裂时,在特征中随机抽取一部分特征,在抽到的特征中间找到最优解,应用于节点进行分裂。

随机森林的方法基于集成的思想,集成了多棵决策树,并对样本和特征进行采样从而避免过拟合。前文随机过采样方法可能引起的过拟合,将在随机森林的分类器部分得到缓解。本文对融合后的特征进行归一化预处理,降低量纲对模型训练的影响,并利用随机过采样方法对小样本类进行过采样,降低数据不平衡性。随机过采样后得到的平衡数据集输入随机森林模型进行训练,其中训练集占比 80%,测试集占比 20%。具体如算法 1 所示。

算法 1 街区品质评估算法

输入:街区卫星图像 picture, 街区非图像型特征表示 Jiequ;

输出:基于数据融合的街区特征表示 Fusion, 各项特征的重要性, 街区品质的预测类别;

```
for picture ∈ {picture1, picture2, ..., picturen}
```

```
  for pixel ∈ all pixels
```

```
    计算属于街区的像素数
```

```
    计算属于绿化的像素数
```

```
  end for
```

```
  计算绿化占比(绿化像素数/街区像素数)
```

```
end for
```

```
for Jiequ ∈ {Jiequ1, Jiequ2, ..., Jiequn}
```

将样本的 picture 和 Jiequ 特征首尾相连生成融合特征 Fusion

```
  构建新数据集
```

```
end for
```

```
for n ∈ [少数类样本个数, 多数类样本个数]
```

```
  从少数类中有放回地随机抽取样本
  构建平衡数据集
```

```
end for
```

```
for number ∈ [1, 100]
```

```
  随机划分训练集和测试集
```

```
  训练集归一化
```

```
  由训练集训练随机森林分类器
```

```
  由随机森林对测试集进行预测
```

```
end for
```

3 验证实验

3.1 实验设置

以北京市西城区展览馆路街道内的各公共空间街区^[30,31]为研究对象(图 2),进行城市街区品质评估方法的实验验证。



图 2 展览馆路街道地块

Fig. 2 Plots of Exhibition Hall Road street

实验使用的真实数据集由 5 位建筑学专家通过实地调研和考察,按少数服从多数原则确定各街区最终的空间品质等级,将各个街区按空间品质分为 A、B、C 3 个等级。Jiequ 数据集包含 141 个数据,每个数据代表一个公共空间街区,各品质类别样本个数及占比如表 1 所示。

表 1 Jiequ 数据集组成

Table 1 Jiequ data set composition

品质类别 Quality category	样本个数 Number of samplers	占比(%) Proportion (%)
A	15	10.64
B	86	60.99
C	40	28.37

本文使用网格法最终确定的随机森林参数为 $n_{\text{estimators}} = 21$, $\text{max_depth} = 7$, $\text{max_features} = 4$, $\text{min_samples_leaf} = 1$, $\text{min_samples_split} = 2$, $\text{criterion} = \text{'gini'}$, $n_{\text{jobs}} = -1$ 。

3.2 结果与分析

3.2.1 使用图像和非图像数据进行街区品质评估结果对比

为了验证图像与融合特征对街区品质特征的表达能力是否更强,且是否有利于提高街区品质评估的正确率,分别使用图像特征和非图像特征,与本文提出的融合特征进行对比。该对比方法可以验证融合特征的可行性和有效性。一方面,本文使用梯度直方图(Histogram of Oriented Gradient, HOG)和像素筛选作为图像特征提取方法;另一方面,本文已将各项测度指标作为数值向量的各维度,可直接作为非图像特征进行后续计算,因此无需进行非图像特征提取。

首先,本文仅使用非图像特征进行街区品质评估,分别选取传统机器学习中较常用的算法:Extreme Gradient Boosting(XGBoost)、逻辑回归(Logistic Regression, LR)、随机森林(RF)和支持向量机(SVM)进行实验对比,结果如表2所示,RF和SVM的正确率较XGBoost和LR提高约7个百分点。

表2 各机器学习模型对比结果

Table 2 Comparison results of machine learning models

算法 Algorithm	正确率(%) Accuracy (%)
XGBoost	75.86
LR	75.86
RF	82.76
SVM	82.76

为探究图像特征提取方法对街区品质评估正确率的影响,本文选择SVM作为分类器,并设置balanced参数,以此减少数据平衡性等其他因素的影响。在不进行图像特征提取的情况下,使用原图像进行街区品质评估的正确率约为42.5%,使用HOG传统特征提取方法进行街区品质评估的正确率约为58%,提高约16个百分点;使用像素筛选法提取图像中的绿色像素(绿色通道值大于其他两通道),并进行街区品质评估的正确率约为63.5%,比原图像法的正确率提高约21个百分点,比HOG方法正确率提高约5个百分点。

由以上两组实验结果可知,使用图像和非图像数据进行街区品质评估的正确率不同,使用图像型数据进行评估的正确率相比非图像型数据下降约20个百分点。因为非图像型数据的各项指标经过专家筛选,所以其街区品质表达能力更强,与街区品质类别的相关性更强,但由于非图像型特征不含空间关系等弊

端,正确率仅为82.76%。因此,基于图像特征和非图像特征在街区品质表达角度上的差异及其互补性,提出将两种特征融合的街区品质评估算法。由于街区数据集的不平衡性,本文使用不同的过采样方法进行对比,实验结果如表3所示。

表3 过采样方法的对比结果

Table 3 Comparison results with the oversampling method

过采样方法 Oversampling method	分类器 Classifier	正确率(%) Accuracy (%)
SMOTE	SVM	84.23
	RF	87.12
Borderline-SMOTE	SVM	57.89
	RF	64.56
ADASYN	SVM	83.57
	RF	87.59
Naive random oversampling	SVM	87.06
	RF	90.98

由表3可见,基于朴素随机过采样方法的街区品质评估正确率高于SMOTE、Borderline-SMOTE和ADASYN方法,且在使用朴素随机过采样方法的前提下,以RF作为分类器的街区品质评估正确率为90.98%,比以SVM作为分类器的正确率提高约4个百分点。可见,本文提出的使用朴素随机过采样方法将融合特征平衡化,并使用RF作为分类器的街区品质评估方法具有可行性。

3.2.2 使用混淆矩阵评价本文方法

从图3的混淆矩阵可以看出,随机生成训练集和测试集后,街区品质评估模型的正确率保持在17,14和16,但对于品质类别B和C的分类效果较差,容易混淆B类和C类而造成分类错误,说明真实样本中B类样本和C类样本具有一定的相似性,甚至在多维特征空间中两种类别的空间存在交集。

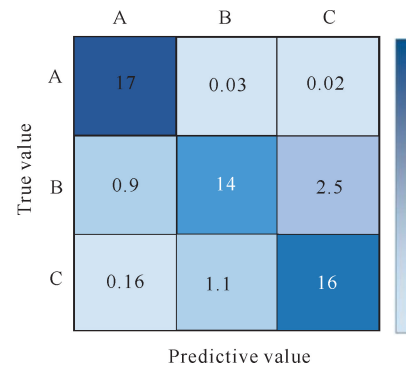


图3 本文方法中各品质类别的混淆矩阵

Fig. 3 Confusion matrix of each quality category in this method

3.2.3 各项指标的重要性评估

本文使用 RF 模型输出各项指标的重要性并排序,重要性从高到低依次为绿化覆盖率、空间视线吸引力、绿化占比、天空宽阔度、人行道宽度、道路交通可达性、植被丰富度、沿街开敞空间可达性、重要历史建筑、街道建筑高宽比、人流聚集度、公交地铁距离、停车干扰度、功能混合度、街道交通潜力、功能密度、停车位数量、公共交通步行可达性、沿街建筑功能种类、小型服务设施种类、公共厕所设施可达性、建筑风貌和出入口数量。由图像像素计算得出的绿化占比的重要性排在第 3,可见其街区品质的表达能力较强,在一定程度上提高了街区品质评估的正确率。重要性排序的结果说明以卫星图像提取的绿化占比作为图像特征,有较大的重要性,其重要性仅次于绿化覆盖率和空间视线吸引力。可见,街区品质与绿化和人类视野感受的关联性较大,本文提出利用卫星图像获取图像特征并进行特征融合的方法是合理且有效的。

综上所述,本文提出的卫星图像代替街景图像进行城市街区绿化情况的特征表示方法,不仅符合城市规划与设计领域对于街区品质评估在概念上的要求,而且确实有利于提高城市街区品质评估模型的性能。随机过采样与随机森林的结合,在本文研究的展览馆路街道内街区数据集上达到了 90.98% 的正确率,相较未使用过采样和多特征融合的方法提高了约 8 个百分点,表明本文提出的算法具有较好的实用性和泛化性。

4 结论

本算法以卫星图像取代街景图像作为输入图像的新方式,经实验证实可有效降低街景图采集位置与视角不统一造成的偏差;针对已有研究注重指标中的某一方面而忽略全面评估的情况,提出将非图像型数据与图像型数据提取的特征进行融合,综合多种数据形式对城市街区各项与品质有关的因素进行全面表征;使用朴素随机过采样方法削弱真实数据集的不平衡性对模型训练的影响,同时使用随机森林作为分类器,利用随机森林能有效降低过拟合的可能性,中和朴素随机过采样容易过拟合的问题。

未来的工作将进一步对图像特征进行挖掘,在数值特征和图像特征数量增加的情况下,提出更有效的多模态特征融合方法。

参考文献

- [1] 龙瀛,唐婧娴.城市街道空间品质大规模量化测度研究进展[J].城市规划,2019,43(6):107-114.
- [2] 李诗卉,杨卓,梁潇,等.东四历史街区:基于多时相街景图片的街道空间品质测度[J].北京规划建设,2016(6):39-48.
- [3] 李智,龙瀛.基于动态街景图片识别的收缩城市街道空间品质变化分析——以齐齐哈尔为例[J].城市建筑,2018(6):21-25.
- [4] 唐婧娴,龙瀛.特大城市中心区街道空间品质的测度:以北京二三环和上海内环为例[J].规划师,2017,33(2):68-73.
- [5] TANG J X, LONG Y. Measuring visual quality of street space and its temporal variation: Methodology and its application in the Hutong area in Beijing [J]. Landscape and Urban Planning, 2019, 191: 103436. DOI: 10. 1016/ j. landurbplan. 2018. 09. 015.
- [6] ASPINALL P, MAVROS P, COYNE R, et al. The urban brain: Analysing outdoor physical activity with mobile EEG [J]. British Journal of Sports Medicine, 2015, 49(4): 272-276.
- [7] RUNDLE A G, BADER M D M, RICHARDS C A, et al. Using google street view to audit neighborhood environments [J]. American Journal of Preventive Medicine, 2011, 40(1): 94-100.
- [8] NAIK N, PHILIPOOM J, RASKAR R, et al. Streets core: Predicting the perceived safety of one million streets capes [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshop. Washington, DC, USA: IEEE, 2014: 793-799.
- [9] EWING R, CERVERO R. Travel and the built environment: A meta-analysis [J]. Journal of the American Planning Association, 2010, 76(3): 265-294.
- [10] 韩君伟.步行街道景观视觉评价研究[D].成都:西南交通大学,2018.
- [11] 龙瀛,周垠.街道活力的量化评价及影响因素分析——以成都为例[J].新建筑,2016(1):52-57.
- [12] 樊钧,唐皓明,叶宇.街道慢行品质的多维度评价与导控策略——基于多源城市数据的整合分析[J].规划师,2019,35(14):5-11.
- [13] 徐磊青,康琦.商业街的空间与界面特征对步行者停留活动的影响——以上海市南京西路为例[J].城市规划学刊,2014(3):104-111.
- [14] 王德,卢银桃,朱玮,等.社区日常服务设施可步行性评价系统开发与应用[J].同济大学学报(自然科学版),

- 2015, 43(12):1815-1822.
- [15] SHEN Y, KARIMI K. Urban function connectivity: Characterisation of functional urban streets with social media check-in data [J]. *Cities*, 2016, 55:9-21.
- [16] 龙瀛, 周垠. 街道活力的量化评价及影响因素分析——以成都为例[J]. *新建筑*, 2016(1):52-57.
- [17] GONG F Y, ZENG Z C, ZHANG F, et al. Mapping sky, tree, and building view factors of street canyons in a high-density urban environment [J]. *Building and Environment*, 2018, 134:155-167.
- [18] 刘伦, 王辉. 城市研究中的计算机视觉应用进展与展望[J]. *城市规划*, 2019, 43(1):117-124.
- [19] 唐婧娴, 龙瀛, 翟炜, 等. 街道空间品质的测度、变化评价与影响因素识别: 基于大规模多时相街景图片的分析[J]. *新建筑*, 2016(5):110-115.
- [20] SERRA M, HILLIER B, KARIMI, K. Exploring countrywide spatial systems: Spatio-structural correlates at the regional and national scales [C]//*Proceedings of the 10th International Space Syntax Symposium (SSS10)*. London, UK: [s. n.], 2015.
- [21] 杨滔. 网络聚集的厚度——区域物质空间形态分析初探[J]. *城市设计*, 2016(5):56-67.
- [22] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: Synthetic minority over-sampling technique [J]. *Journal of Artificial Intelligence Research*, 2002, 16(1):321-357.
- [23] HUI H, WANG W Y, MAO B H. Borderline SMOTE: A new over-sampling method in imbalanced data sets learning [C]//*Proceedings of the 2005 international conference on Advances in Intelligent Computing - Volume Part I*. Hefei, China: Springer, 2005: 878-887.
- [24] HE H B, BAI Y, GARCIA E A, et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning [C]//*2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. Hong Kong, China: IEEE, 2008.
- [25] 李苗裔, 杨忠豪, 薛峰. 基于多源数据的城市街道绿化品质测度与规划设计提升策略——以福州主城区为例[J]. *风景园林*, 2021, 28(2):62-68.
- [26] 莫怡晨, 李飏. 基于多数据源的城市空间数据库设计研究——以意大利普拉托为例[C]//*2020年全国建筑院系建筑数字技术教学与研究学术研讨会论文集*. 重庆: 全国高等学校建筑学专业指导委员, 2020:183-189.
- [27] 孙瑞丰, 王语嫣. 城市街道空间品质评价体系构建研究[J]. *安徽建筑*, 2021, 28(3):16-17.
- [28] 叶洋, 费腾. 基于街景图像的街道空间步行品质评价研究[J]. *当代建筑*, 2020(11):123-125.
- [29] 刘斯霞. 基于多源数据的城市街道空间品质评估[J]. *城市建筑*, 2020, 17(18):14-16.
- [30] 李美艳. 基于空间句法的城市街道空间研究——以北京市展览馆路内街道为例[J]. *中国建筑装饰装修*, 2021(9):128-129.
- [31] 李美艳, 张润萌, 欧阳文. 商业POI布局和城市街区空间结构的相关性研究——以北京西城区展览馆路街区为例[J]. *北京规划建设*, 2021(5):125-127.

Block Quality Evaluation Based on Oversampling and Satellite Image Multi-feature Fusion

GUO Maozu^{1,2}, WANG Sijia^{1,2}, WANG Pengyue^{2,3}, LI Yang^{1,2}, ZHAO Lingling⁴

(1. College of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China; 2. Beijing Key Laboratory of Intelligent Processing for Building Big Data (Beijing University of Civil Engineering and Architecture), Beijing, 100044, China; 3. College of Architecture and Urban Planning, Beijing University of Civil Engineering and Architecture, Beijing, 100044, China; 4. College of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, 150001, China)

Abstract: The quality evaluation of urban blocks is one of the important steps in urban planning. In order to solve the problem of image content deviation caused by the inconsistent collection position and angle of street

view image used in the existing block quality evaluation model, this article proposes to use the satellite image with unified perspective and easy access to replace the street view image. At the same time, considering a variety of block quality measurement indicators, the characteristics of non image data and image data are fused, and the quality of urban blocks is expressed from multiple perspectives. Then, for small samples and data imbalance scenarios, combined with random oversampling and Random Forest (RF) in machine learning, the quality of urban blocks is evaluated based on the fusion data. The experimental results show that this research method can effectively carry out a comprehensive urban block quality expression and effectively reduce over fitting. Compared with the methods without over sampling and multi-feature fusion, the accuracy rate is increased by about 8 percentage points.

Key words: quality evaluation of urban blocks; feature fusion; simple random oversampling; random forest; satellite image

责任编辑:陆 雁



微信公众号投稿更便捷

联系电话:0771-2503923

邮箱:gxxk@gxas.cn

投稿系统网址:<http://gxxk.ijournal.cn/gxxk/ch>